

# Evolutionary history of regulatory variation in human populations

Tuuli Lappalainen and Emmanouil T. Dermitzakis\*

Department of Genetic Medicine and Development, University of Geneva Medical School, 1411-4 Geneva, Switzerland

Received August 5, 2010; Revised and Accepted September 15, 2010

**Genetic variation in the regulation of gene expression is likely to be a major contributor to phenotypic variation in humans, and it also constitutes an important target of recent natural selection in human populations and plays a major role in morphological evolution. The increasing amount of data of genome and transcriptome variation is now leading to a better annotation of regulatory elements and a growing understanding of how the evolution of gene regulation has shaped human diversity. In this review, we discuss the evolutionary history of the variation in the expression of protein-coding genes in humans. We outline the current methodology for mapping regulatory variants and their distribution in human populations. General mechanisms of regulatory evolution are discussed with a special emphasis on different selective processes targeting gene regulation in humans.**

## INTRODUCTION

Analysis of regulatory variation has been motivated by a quest for understanding the sources of phenotypic variation in humans, including variation in susceptibility to disease. Genetic differences in the regulation of gene expression may also underlie some of evolutionary adaptive phenotypic differences between human populations, and from a longer evolutionary perspective, the evolution of human-specific traits that distinguish us from other primates has been a major focus of research. The importance of gene regulation in morphological evolution has been acknowledged and debated for decades (1–7). Recently, the field of evolutionary genetics has witnessed an accumulation of evidence of regulatory changes underlying phenotypic differences within and between species—first through case examples, but increasingly through genome-wide analysis of genomes and transcriptomes. Now, we are increasing our understanding of how the information in the genetic code is transferred to the transcriptome, to proteins, and thereon to phenotypes at the cellular, systemic and organismal levels. We are learning how different types of genetic variations alter these pathways, and how variants of different functional categories are being shuffled by the evolutionary process. In this review article, we will discuss population genetics of regulatory variation affecting the expression of protein-coding genes in humans,

and the evolutionary history of this variation. The non-coding part of the transcriptome and its evolution has been discussed elsewhere (8–12).

## ANALYSING GENOMES AND TRANSCRIPTOMES

The analysis of regulatory variation requires data of both genetic and transcriptome data. Levels of gene expression have been analyzed now for almost 10 years in a genome-wide manner by expression arrays, and genome-wide analysis of human genetic variation was made possible about 5 years ago through the development of array-based genotyping of hundreds of thousands of single nucleotide polymorphisms (SNPs).

Despite the wide variety of approaches relying on these techniques, the recent advance in sequencing technologies has opened a range of new exciting possibilities for the analysis of the genome and its function. Sequencing of mRNA offers a much more accurate analysis of splicing patterns, expression levels and allele-specific expression than array-based technologies (13–18). Transcriptome sequencing has also revolutionized the comparison of gene expression patterns between species by eliminating the need to rely on pre-designed probes that have been available only for a limited set of species with established genome annotation.

\*To whom correspondence should be addressed. Tel: +41 22 379 5483; Fax: +41 22 379 5706; Email: emmanouil.dermitzakis@unige.ch

The analysis of genetic variation is also shifting from genotyping of pre-selected SNPs to genomic re-sequencing, which offers not only a denser coverage of variants and accurate genotyping of structural variation but also a more even coverage of the frequency spectrum (17) ([www.1000genomes.org](http://www.1000genomes.org)) without ascertainment bias towards well-studied populations in array SNP selection, which has been a concern in population genetic studies (19). The new technologies are also being used for *de novo* sequencing as well for population-based re-sequencing of non-human species to discover genetic variation in other organisms (20,21) ([www.sanger.ac.uk/modelorgs/mousegenomes](http://www.sanger.ac.uk/modelorgs/mousegenomes)).

Furthermore, an increasing understanding of the mechanisms of genome function and its evolution is being gained through sequencing applications for assaying, for example, transcription factor binding, methylation patterns and chromatin structure. Integrating these data into knowledge of genetic and transcriptome analysis will shed light on regulatory networks and the annotation of regulatory elements (12,17,22–26).

## MAPPING REGULATORY VARIATION

Several approaches have been developed to find genetic variants that affect gene expression. The most common method has been testing for association between the genotype classes of a genetic variant and gene expression levels to map expression quantitative trait loci (eQTLs), mostly in *cis* close to the target gene, but also in *trans* (13,14,27–35). eQTL analysis captures only common regulatory variation, because statistical power to detect association to expression levels decreases sharply with minor allele frequency. Another approach has been to study allele-specific expression, where allelic imbalance in the mRNA production between coding heterozygous polymorphisms is used as a signal of *cis*-regulatory variation (36–40). This method has its own limitations and sources of error, but has better power to find rare regulatory variants in *cis*, accessible especially in RNA-sequencing data (13,14,18). Altogether, these studies have shed light on the general patterns of regulatory variation in human populations and provided the tools for further studies of the role of regulatory variation, for example, in human disease and evolution.

These approaches do not, however, give direct information of the causal variants that alter gene expression, because typically a large number of variants show significant association to expression values of a gene because of the linkage disequilibrium (LD) in the human genome, i.e. the strong correlation of genetic variants located up to tens or even hundreds of kilobases from each other. This makes defining the causal variant difficult even when full information of all the genetic variation is available through genomic resequencing (Montgomery *et al.*, in preparation). Figure 1 illustrates an example of this: SNPs from array (Fig. 1A) as well as resequencing (Fig. 1B) data sets have several markers that show significant association due to being in LD with each other. However, the peak region is easier to distinguish in the denser resequencing data, especially from the African population where the extent of LD is lower than in the Europeans due to differences in

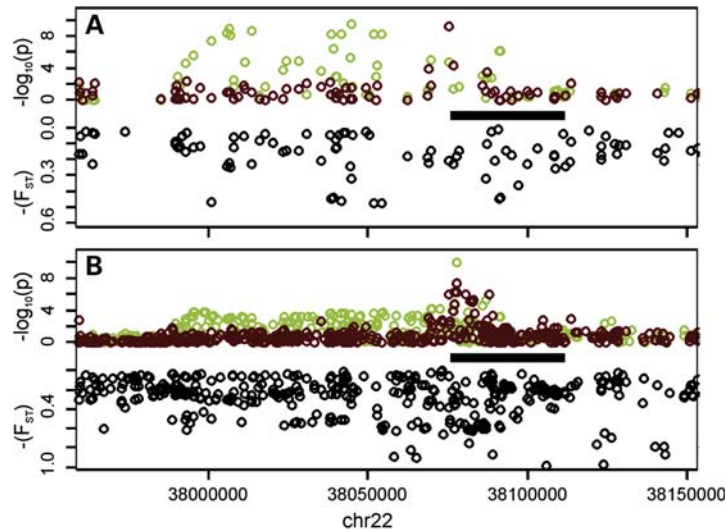
population history. Additional information of the location of the variants relative to functional elements such as the transcription start site, transcription factor binding sites and splice sites, as well as evolutionary conservation and differences between populations can also be used to model the most likely causal variant (41).

Many types of mutation can affect gene regulation and lead to variation in expression patterns within and between species. Even though SNPs have been studied most, structural changes of different sizes are likely to contribute as well: in human populations, 20% of eQTLs appear to be caused by large copy number variations typically of >100 kb in length (29). However, this proportion is likely to rise when data from small insertions and deletions, obtained from resequencing data, are added to the analysis—although at the same time, this will complicate the inference of causality (Montgomery *et al.*, in preparation). The importance of structural changes has been observed in interspecies comparisons, too: there is a positive correlation between the density of structural mutations and the extent of expression level changes between species (42,43). The mechanism of how structural variation affects gene expression appears to be independent of a simple dosage effect both within and between species, suggesting that structural variation in proximal regulatory elements is often causing the change in expression patterns (29,42,44).

## PATTERNS OF REGULATORY VARIATION IN HUMAN POPULATIONS

Human populations show significant differences in gene expression levels: in analyses of cell lines from different populations, 17–29% of genes have shown expression differences between European, African and Asian populations (27,30,45), and about 15% of the total expression variation between an European American and an African population could be attributed to differences between the populations (45,46). A large part of this variation is genetic: overall heritability of gene expression in humans has been calculated to be 0.43 by using an admixed population (46), and in different studies expression levels for 13–31% of human genes have been estimated to have heritability over 0.2 (30–33,47,48). Also splicing of exons is known to have genetically determined variation between individuals and populations, although this has only recently become apparent from large-scale studies (13,14,49–52). However, environmental and technical contribution to the observed gene expression variation is not to be overlooked (53,54). In particular, if gene expression is measured not from cell lines but from individuals exposed to different environments, even genetically similar populations have been shown to have large expression differences (55,56), and thus the heritability values obtained from cell lines are likely to be overestimates of the true values in human populations.

Over 1000 eQTLs have been mapped across the human genome in different populations, mostly *cis*-eQTLs, but also some *trans*-eQTLs and a much smaller number of splicing QTLs (13,14,27–35). Studies of allele-specific expression capture some of the same loci but also others, with up to



**Figure 1.** Colocalization of eQTL (shown in red and green) and selection (shown in black) signals in the *SYNGR1* gene, shown for two populations and two SNP densities. (A) Above the x-axis, 1000 genomes low-coverage SNP (from the pilot 3 release, www.1000genomes.org) associations to array expression values of *SYNGR1* (30) in a population of European background (CEU) shown in green ( $n = 55$ ), and Yoruba from Ibadan, Nigeria (YRI) in red ( $n = 53$ ). Below the x-axis are shown overall  $F_{ST}$  values that measure allele frequency differences between CEU, YRI, Chinese from Beijing (CHB) and Japanese from Tokyo (JPT). (B) On the upper hand side, HapMap3 SNP associations (www.hapmap.org) to the expression levels of *SYNGR1* in CEU (green;  $n = 109$ ) and YRI (red;  $n = 108$ ) (Stranger *et al.*, in preparation). Corresponding signals were observed in other six HapMap3 populations. Below the x-axis, overall  $F_{ST}$  between all the 11 populations of the HapMap3 data set. The eQTL analysis has been done using Spearman rank correlation (30) only for common SNPs (minor allele frequency  $> 5\%$ ); the x-axis shows the  $-\log_{10}$  of the  $P$ -value of the correlation. The black bar shows the location of the *SYNGR1* gene which encodes for a membrane protein in presynaptic vesicles. High allele frequency differentiation, measured here by  $F_{ST}$ , is a classical signal of natural selection (94), and in this gene, another test of recent positive selection, the haplotype-based  $iHS$ , also overlaps with the eQTL signal (78), rendering further support that the expression of this gene may be a target of recent natural selection. The higher LD in Europeans is clearly visible in the long range of SNPs with a significant association signal compared with the relatively narrow peak in the Yoruba, but distinguishing the causal variant even from the 1000 genomes data remains difficult. Even though the landscape of the eQTL signal is clearer in the resequencing data, the SNP array data of HapMap3 shows essentially the same pattern of association. The better  $P$ -values in the HapMap3 data for some of the shared SNPs are due to the higher sample size.

30% of genes having signs of common regulatory variation in *cis* (36–38,40). Analyses of different populations generally show a significant overlap with about one-third of eQTLs being shared between populations from Africa, Europe and Asia (30). In general, population differences in gene expression and eQTL sharing appear to follow genetic differentiation across the genome as well as in a genome-wide scale across a large set of populations from different continents (Stranger *et al.*, in preparation, 27,46). Shared eQTLs appear to have nearly always the same direction of effect in different populations as well as similar fold changes, suggesting that while the extent to which individual eQTLs affect gene expression depend on allele frequencies that vary between populations, the underlying regulatory mechanisms are shared (Stranger *et al.*, in preparation, 27,30).

Mapping of regulatory variation in *cis* has been much easier than in *trans* due to higher effect sizes and easier control of multiple testing problems (30,31,33,47,57), and thus patterns of *cis*-variation are much better understood. However, regulatory variation in *cis* probably represents a minority of the total heritable proportion of variation in gene expression levels: estimates derived from an analysis of an admixed human population suggests that it accounts for only 12% of the total variation (46), which is consistent with a higher contribution of *trans*-variation observed in *Drosophila* (58), although defining *cis*- and *trans*-variation is often difficult and varies between studies (57,59,60). Most human studies have relied on transformed lymphoblastoid cell lines, but it

is now known that regulatory differences between different cell types are much bigger than between populations, with common tissue-specific effects, where an eQTL affects gene expression only in a part of the tissues where the gene is expressed (32,49,61,62).

## MECHANISMS OF REGULATORY EVOLUTION

The relative contribution of regulatory and coding changes in evolutionary change is one of the big open questions in evolutionary genetics. In general, compared with changes in the amino acid sequences, the functional consequences of regulatory mutations, especially in *cis*, are thought to provide more flexible material for natural selection with fewer functional trade-offs: while a non-synonymous mutation generally leads to a qualitative change in protein structure wherever the gene is expressed, regulatory changes are quantitative, and may often be activated only in particular environmental conditions, developmental stages or tissues. Furthermore, differences in patterns of selection may arise from *cis*-regulatory variants being usually codominant, whereas coding and *trans*-regulatory variants are more often recessive (31,60,63). However, despite many comparisons of the evolution of regulatory and coding regions, no consensus of their relative importance has been reached (2–4,6,7,53,64–67). Unbiased analysis is challenging due to the difficulty of predicting functional consequences of non-coding mutations, and an additional complication arises from LD between coding and

flanking regions, which may lead to correlated patterns even in the absence of similar selective effects. Interaction between *cis*-regulatory and coding non-synonymous variants is also not to be ignored as the expression of the two alleles of coding variants is often imbalanced (36,68), but the phenotypic and evolutionary consequences of such interactions are still poorly known. Furthermore, several studies suggest that natural selection has targeted *cis*- and *trans*-regulatory elements differently: *cis*-elements appear to be more frequent targets of positive selection and contribute more to interspecies differences in gene expression compared with more-constrained *trans*-variation (58,63,65,69)—possibly because *cis*-elements have more tissue-specific effects and are thus less likely to have pleiotropic effects across tissues than *trans*-variants.

## REGULATORY VARIATION AS A TARGET OF NATURAL SELECTION

Although gene expression has been suggested to be under less strict purifying selection than coding variation (1), and even selective neutrality has been proposed (66), gene expression levels both between and within populations often show evolutionary constraint (15,53,67,70). Regulatory regions of the genome are often strongly conserved between species, and have a lower degree of variation also within species (12,71–74). In human populations, SNPs in 5' and 3' regions show an enrichment of low differentiation between populations suggesting purifying selection (75). Also, variation in transcription factor binding between humans and between human and chimpanzee is much more common in sites far away from transcription start sites of genes, suggesting that natural selection eliminates variation that would change expression patterns of genes (22).

Because of the lack of power to detect rare eQTLs, the genes regulated by known eQTLs generally tolerate common regulatory variation in humans. Thus, their regulation is likely to be less constrained, and it has been shown that the carriers of the ancestral haplotype of common human eQTLs do not show expression levels closer to the chimpanzees, suggesting low constraint with several rounds of regulatory variants fixating in both species (Montgomery and Dermitzakis, unpublished data). This is supported by *cis*-variation being rarer in gene-dense regions that are likely to be more constrained (69), and by the significant overlap between human and mouse genes that show allele-specific expression, which suggests low constraint in these genes in mammals (76). Thus, eQTL data are inherently biased towards common variants in genes whose expression is not heavily constrained, but the novel possibilities for mapping rare regulatory variation from RNA sequencing data (13) will now yield a more complete catalogue of different regulatory variants and enable more comprehensive analyses of evolutionary processes affecting human gene expression.

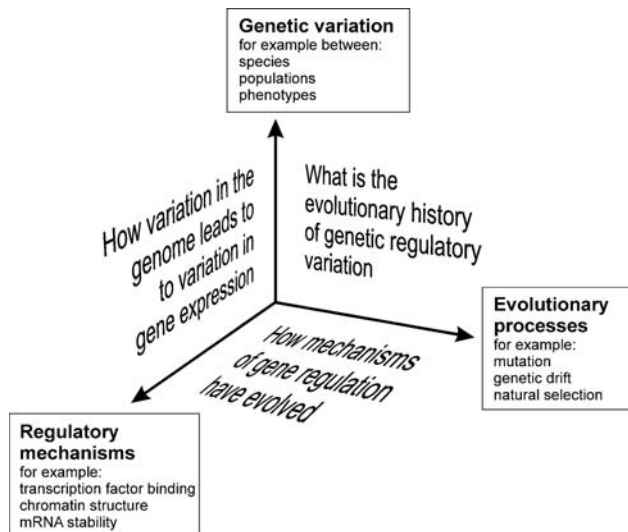
Even though a combination of purifying selection and selective neutrality probably predominates the evolution of regulatory and other types of functional variation, gene regulation

constitutes a potential target for adaptive evolution, and eQTLs appear to be enriched for signs of recent positive selection in humans (77). Figure 1 provides an example of such a gene: the SNPs with the highest association to gene expression levels also have large allele frequency differences between populations, which is often a sign of recent positive selection that drives a beneficial allele to a higher frequency. Also 5' regions that often harbor regulatory elements appear enriched for signs of positive selection (64,75). There are several case examples of recent adaptations through regulatory variation, for instance the convergent emergence of lactase tolerance via continued expression of the *LCT* gene after childhood (78), and resistance to malaria through a tissue-specific inactivation of the Duffy antigen (79,80). Macro-evolutionary differences between species may also derive from regulatory changes, and much work has been dedicated to characterizing events of positive selection that underlie human-specific adaptations. A large number of genes show a change in the expression level in humans compared with other primates, which may sometimes be a sign of directional selection (67). Many well-characterized regulatory adaptations are already known, such as the rapid evolution in the *HACNS1* enhancer that may have contributed to limb development in the human lineage (81,82).

Whether the regulation of genes that are expressed across multiple tissues are more or less frequent targets of selection than tissue-specific ones is a matter of debate, with results suggesting more regulatory constraint on ubiquitously expressed genes (66,67,83), or less constraint (84), as well as little overall correlation (64). Evolution of the human brain has been a topic of particular interest, and several studies have found an enrichment of regulatory adaptation in the brain (64,66,67). Additionally, the evolution of male reproduction, the brain and the dietary system appears to have been dominated by regulatory change, as opposed to an enrichment of coding changes in the evolution of, for instance, the immunological system (64,85). To date, little is known of the systemic targets of recent positive selection in humans due to the unavailability of expression data from multiple tissues from a variety of populations. In general, the modest degree of overlap between different studies suggests that the analyses of systemic targets of regulatory evolution have not been very robust. While straightforward grouping of genes according to tissue of expression has provided a good starting point for understanding the evolution of gene expression, it lacks the resolution that, for example, network-based approaches may have (86).

Changes in gene expression are likely to be one of the major mechanisms underlying differences in susceptibility to disease between individuals, for both Mendelian and common disease (31–33,36,87–90). Disease-associated genes have been observed to be enriched for negative selection both in the coding and regulatory regions, but there are also several examples of disease genes with signs of positive selection in their regulatory regions (64,77,91–93). However, it remains mostly unknown how often disease-causing coding and regulatory genetic variation has arisen through inefficiency of purifying selection, and to what degree it is a result of evolutionary trade-offs or past positive selection.





**Figure 2.** Multiple dimensions of the analysis of the evolutionary history of regulatory variation.

## CONCLUSIONS AND FUTURE PROSPECTS

The effects of coding variation have been studied for decades on several levels, ranging from the cell to tissues and organisms, and further to the role of coding variation in disease and evolution. Now, this spectrum is being studied also for regulatory variation. Although mapping the causal variants of *cis*-regulatory variation in the eQTL region still remains a challenge due to the correlation structure of genetic variants, the annotation of proximal regulatory elements has progressed rapidly, and will become even more accurate through the many applications of novel sequencing technologies. Thus, finding loci with regulatory variants of high effect sizes in *cis* is starting to become straightforward. However, this likely accounts only for a small proportion of all the heritable variation of gene expression, and the biggest challenge now lies in understanding the contributions of common regulatory polymorphisms in *trans*, and rare regulatory variants. The different layers of transcriptional regulation and the complexity of feedback networks are difficult to untangle, given the statistical challenge of testing interactions across the entire genome. Furthermore, many genetic regulatory effects are likely not stable and ubiquitous, but are mediated through modifications of regulatory networks in a cell-type specific manner during different developmental stages or as a response to particular environmental conditions. Yet, information of all of the aspects of the regulatory landscape is essential if we are to understand how variation in the genomes has given rise to the biological complexity we observe around us, within and between species (Fig. 2).

Evolutionary analysis of gene regulation has now moved from case examples to genome-wide approaches. Studies of individual genes and regulatory elements have provided intriguing examples of processes of evolutionary adaptation, but it is unfeasible to collect sufficient experimentally validated case examples to get an unbiased view of general evolutionary mechanisms and their relative importance. Genome-wide scanning is another approach to find targets of natural selection. However, it often yields long lists of candidate genes

whose validation is difficult and that are often biased towards specific types and ages of selection. The search for general trends behind these gene lists has often been based on categorization according to gene ontology or tissue of expression, but these approaches lack resolution and have often yielded relatively inconsistent results between different studies. A general problem in evolutionary genetics is that it is relatively easy to come up with attractive stories of possible adaptive mechanisms even in the absence of real evidence. Some degree of uncertainty is inevitable because the evolutionary history cannot be rerun to obtain a truly independent replication, but especially now in the era of massive genomic data sets, we must aim to design and conduct studies that test well-defined hypotheses and answer specific questions about the evolution of genomes.

*Conflict of Interest statement.* None declared.

## FUNDING

The funding was provided by the Louis-Jeantet foundation, the Academy of Finland and the Emil Aaltonen foundation.

## REFERENCES

- Wray, G.A. (2007) The evolutionary significance of *cis*-regulatory mutations. *Nat. Rev. Genet.*, **8**, 206–216.
- Carroll, S.B. (2008) Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell*, **134**, 25–36.
- Lynch, V.J. and Wagner, G.P. (2008) Resurrecting the role of transcription factor change in developmental evolution. *Evolution*, **62**, 2131–2154.
- Prud'homme, B., Gompel, N. and Carroll, S.B. (2007) Emerging principles of regulatory evolution. *Proc. Natl Acad. Sci. USA*, **104**(Suppl. 1), 8605–8612.
- King, M.C. and Wilson, A.C. (1975) Evolution at two levels in humans and chimpanzees. *Science*, **188**, 107–116.
- Hoekstra, H.E. and Coyne, J.A. (2007) The locus of evolution: evo devo and the genetics of adaptation. *Evolution*, **61**, 995–1016.
- Carroll, S.B. (2005) Evolution at two levels: on genes and form. *PLoS Biol.*, **3**, e245.
- Mattick, J.S. (2009) The genetic signatures of noncoding RNAs. *PLoS Genet.*, **5**, e1000459.
- Chen, K. and Rajewsky, N. (2007) The evolution of gene regulation by transcription factors and microRNAs. *Nat. Rev. Genet.*, **8**, 93–103.
- Ponting, C.P., Oliver, P.L. and Reik, W. (2009) Evolution and functions of long noncoding RNAs. *Cell*, **136**, 629–641.
- Pheasant, M. and Mattick, J.S. (2007) Raising the estimate of functional human sequences. *Genome Res.*, **17**, 1245–1253.
- ENCODE Project Consortium, Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T. *et al.* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **447**, 799–816.
- Montgomery, S.B., Sammeth, M., Gutierrez-Arcelus, M., Lach, R.P., Ingle, C., Nisbett, J., Guigo, R. and Dermitzakis, E.T. (2010) Transcriptome genetics using second generation sequencing in a caucasian population. *Nature*, **464**, 773–777.
- Pickrell, J.K., Marioni, J.C., Pai, A.A., Degner, J.F., Engelhardt, B.E., Nkadori, E., Veyrieras, J.B., Stephens, M., Gilad, Y. and Pritchard, J.K. (2010) Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature*, **464**, 768–772.
- Blekhan, R., Marioni, J.C., Zumbo, P., Stephens, M. and Gilad, Y. (2010) Sex-specific and lineage-specific alternative splicing in primates. *Genome Res.*, **20**, 180–189.

16. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat. Methods*, **5**, 621–628.
17. Mardis, E.R. (2008) Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.*, **9**, 387–402.
18. Pastinen, T. (2010) Genome-wide allele-specific analysis: insights into regulatory variation. *Nat. Rev. Genet.*, **11**, 533–538.
19. Albrechtsen, A., Nielsen, F.C. and Nielsen, R. (2010) Research article: ascertainment biases in SNP chips affect measures of population divergence. *Mol. Biol. Evol.*, in press.
20. Turner, D.J., Keane, T.M., Sudbery, I. and Adams, D.J. (2009) Next-generation sequencing of vertebrate experimental organisms. *Mamm. Genome*, **20**, 327–338.
21. Genome 10K Community of Scientists. (2009) Genome 10K: a proposal to obtain whole-genome sequence for 10,000 vertebrate species. *J. Hered.*, **100**, 659–674.
22. Kasowski, M., Grubert, F., Heffelfinger, C., Hariharan, M., Asabere, A., Waszak, S.M., Habegger, L., Rozowsky, J., Shi, M., Urban, A.E. *et al.* (2010) Variation in transcription factor binding among humans. *Science*, **328**, 232–235.
23. Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B. *et al.* (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, **454**, 766–770.
24. Shoemaker, R., Deng, J., Wang, W. and Zhang, K. (2010) Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. *Genome Res.*, **20**, 883–889.
25. Maynard, N.D., Chen, J., Stuart, R.K., Fan, J.B. and Ren, B. (2008) Genome-wide mapping of allele-specific protein-DNA interactions in human cells. *Nat. Methods*, **5**, 307–309.
26. Morozova, O. and Marra, M.A. (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics*, **92**, 255–264.
27. Spielman, R.S., Bastone, L.A., Burdick, J.T., Morley, M., Ewens, W.J. and Cheung, V.G. (2007) Common genetic variants account for differences in gene expression among ethnic groups. *Nat. Genet.*, **39**, 226–231.
28. Stranger, B.E., Forrest, M.S., Clark, A.G., Minichiello, M.J., Deutsch, S., Lyle, R., Hunt, S., Kahl, B., Antonarakis, S.E., Tavaré, S. *et al.* (2005) Genome-wide associations of gene expression variation in humans. *PLoS Genet.*, **1**, e78.
29. Stranger, B.E., Forrest, M.S., Dunning, M., Ingle, C.E., Beazley, C., Thorne, N., Redon, R., Bird, C.P., de Grassi, A., Lee, C. *et al.* (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*, **315**, 848–853.
30. Stranger, B.E., Nica, A.C., Forrest, M.S., Dimas, A., Bird, C.P., Beazley, C., Ingle, C.E., Dunning, M., Flicek, P., Koller, D. *et al.* (2007) Population genomics of human gene expression. *Nat. Genet.*, **39**, 1217–1224.
31. Dixon, A.L., Liang, L., Moffatt, M.F., Chen, W., Heath, S., Wong, K.C., Taylor, J., Burnett, E., Gut, I., Farrell, M. *et al.* (2007) A genome-wide association study of global gene expression. *Nat. Genet.*, **39**, 1202–1207.
32. Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A.S., Zink, F., Zhu, J., Carlson, S., Helgason, A., Walters, G.B., Gunnarsdottir, S. *et al.* (2008) Genetics of gene expression and its effect on disease. *Nature*, **452**, 423–428.
33. Goring, H.H., Curran, J.E., Johnson, M.P., Dyer, T.D., Charlesworth, J., Cole, S.A., Jowett, J.B., Abraham, L.J., Rainwater, D.L., Comuzzie, A.G. *et al.* (2007) Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat. Genet.*, **39**, 1208–1216.
34. Morley, M., Molony, C.M., Weber, T.M., Devlin, J.L., Ewens, K.G., Spielman, R.S. and Cheung, V.G. (2004) Genetic analysis of genome-wide variation in human gene expression. *Nature*, **430**, 743–747.
35. Cheung, V.G., Spielman, R.S., Ewens, K.G., Weber, T.M., Morley, M. and Burdick, J.T. (2005) Mapping determinants of human gene expression by regional and genome-wide association. *Nature*, **437**, 1365–1369.
36. Ge, B., Pokholok, D.K., Kwan, T., Grundberg, E., Morcos, L., Verlaan, D.J., Le, J., Koka, V., Lam, K.C., Gagne, V. *et al.* (2009) Global patterns of cis variation in human cells revealed by high-density allelic expression analysis. *Nat. Genet.*, **41**, 1216–1222.
37. Verlaan, D.J., Ge, B., Grundberg, E., Hoberman, R., Lam, K.C., Koka, V., Dias, J., Gurd, S., Martin, N.W., Mallmin, H. *et al.* (2009) Targeted screening of cis-regulatory variation in human haplotypes. *Genome Res.*, **19**, 118–127.
38. Pastinen, T., Ge, B., Gurd, S., Gaudin, T., Dore, C., Lemire, M., Lepage, P., Harmsen, E. and Hudson, T.J. (2005) Mapping common regulatory variants to human haplotypes. *Hum. Mol. Genet.*, **14**, 3963–3971.
39. Pant, P.V., Tao, H., Beilharz, E.J., Ballinger, D.G., Cox, D.R. and Frazer, K.A. (2006) Analysis of allelic differential expression in human white blood cells. *Genome Res.*, **16**, 331–339.
40. Serre, D., Gurd, S., Ge, B., Sladek, R., Sinnett, D., Harmsen, E., Bibikova, M., Chudin, E., Barker, D.L., Dickinson, T. *et al.* (2008) Differential allelic expression in the human genome: a robust approach to identify genetic and epigenetic cis-acting mechanisms regulating gene expression. *PLoS Genet.*, **4**, e1000006.
41. Veyrieras, J.B., Kudaravalli, S., Kim, S.Y., Dermitzakis, E.T., Gilad, Y., Stephens, M. and Pritchard, J.K. (2008) High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet.*, **4**, e1000214.
42. Blekhan, R., Oshlack, A. and Gilad, Y. (2009) Segmental duplications contribute to gene expression differences between humans and chimpanzees. *Genetics*, **182**, 627–630.
43. De, S., Teichmann, S.A. and Babu, M.M. (2009) The impact of genomic neighborhood on the evolution of human and chimpanzee transcriptome. *Genome Res.*, **19**, 785–794.
44. Hurles, M.E., Dermitzakis, E.T. and Tyler-Smith, C. (2008) The functional impact of structural variation in humans. *Trends Genet.*, **24**, 238–245.
45. Storey, J.D., Madeoy, J., Strout, J.L., Wurfel, M., Ronald, J. and Akey, J.M. (2007) Gene-expression variation within and among human populations. *Am. J. Hum. Genet.*, **80**, 502–509.
46. Price, A.L., Patterson, N., Hancks, D.C., Myers, S., Reich, D., Cheung, V.G. and Spielman, R.S. (2008) Effects of cis and trans genetic ancestry on gene expression in African Americans. *PLoS Genet.*, **4**, e1000294.
47. Schadt, E.E., Monks, S.A., Drake, T.A., Lusis, A.J., Che, N., Colinao, V., Ruff, T.G., Milligan, S.B., Lamb, J.R., Cavet, G. *et al.* (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature*, **422**, 297–302.
48. Monks, S.A., Leonardson, A., Zhu, H., Cundiff, P., Pietrusiak, P., Edwards, S., Phillips, J.W., Sachs, A. and Schadt, E.E. (2004) Genetic inheritance of gene expression in human cell lines. *Am. J. Hum. Genet.*, **75**, 1094–1105.
49. Wang, E.T., Sandberg, R., Luo, S., Khrebukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P. and Burge, C.B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature*, **456**, 470–476.
50. Zhang, W., Duan, S., Bleibel, W.K., Wisel, S.A., Huang, R.S., Wu, X., He, L., Clark, T.A., Chen, T.X., Schweitzer, A.C. *et al.* (2009) Identification of common genetic variants that account for transcript isoform variation between human populations. *Hum. Genet.*, **125**, 81–93.
51. Kwan, T., Benovoy, D., Dias, C., Gurd, S., Provencher, C., Beaulieu, P., Hudson, T.J., Sladek, R. and Majewski, J. (2008) Genome-wide analysis of transcript isoform variation in humans. *Nat. Genet.*, **40**, 225–231.
52. Fraser, H.B. and Xie, X. (2009) Common polymorphic transcript variation in human disease. *Genome Res.*, **19**, 567–575.
53. Gilad, Y., Rifkin, S.A. and Pritchard, J.K. (2008) Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet.*, **24**, 408–415.
54. Choy, E., Yelensky, R., Bonakdar, S., Plenge, R.M., Saxena, R., De Jager, P.L., Shaw, S.Y., Wolfish, C.S., Slavik, J.M., Cotsapas, C. *et al.* (2008) Genetic analysis of human traits in vitro: drug response and gene expression in lymphoblastoid cell lines. *PLoS Genet.*, **4**, e1000287.
55. Idaghdour, Y., Storey, J.D., Jadallah, S.J. and Gibson, G. (2008) A genome-wide gene expression signature of environmental geography in leukocytes of moroccan amazighs. *PLoS Genet.*, **4**, e1000052.
56. Idaghdour, Y., Czika, W., Shianna, K.V., Lee, S.H., Visscher, P.M., Martin, H.C., Miclaus, K., Jadallah, S.J., Goldstein, D.B., Wolfinger, R.D. *et al.* (2010) Geographical genomics of human leukocyte gene expression variation in southern morocco. *Nat. Genet.*, **42**, 62–67.
57. Petretto, E., Mangion, J., Dickens, N.J., Cook, S.A., Kumaran, M.K., Lu, H., Fischer, J., Maatz, H., Kren, V., Pravenec, M. *et al.* (2006) Heritability and tissue specificity of expression quantitative trait loci. *PLoS Genet.*, **2**, e172.
58. Wittkopp, P.J., Haerum, B.K. and Clark, A.G. (2008) Regulatory changes underlying expression differences within and between drosophila species. *Nat. Genet.*, **40**, 346–350.

59. Wang, H.Y., Fu, Y., McPeck, M.S., Lu, X., Nuzhdin, S., Xu, A., Lu, J., Wu, M.L. and Wu, C.I. (2008) Complex genetic interactions underlying expression differences between drosophila races: analysis of chromosome substitutions. *Proc. Natl Acad. Sci. USA*, **105**, 6362–6367.
60. McManus, C.J., Coolon, J.D., Duff, M.O., Eipper-Mains, J., Graveley, B.R. and Wittkopp, P.J. (2010) Regulatory divergence in drosophila revealed by mRNA-seq. *Genome Res.*, **20**, 816–825.
61. Dimas, A.S., Deutsch, S., Stranger, B.E., Montgomery, S.B., Borel, C., Attar-Cohen, H., Ingle, C., Beazley, C., Gutierrez Arcelus, M., Sekowska, M. *et al.* (2009) Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science*, **325**, 1246–1250.
62. Kwan, T., Grundberg, J., Koka, V., Ge, B., Lam, K.C., Dias, C., Kindmark, A., Mallmin, H., Ljunggren, O., Rivadeneira, F. *et al.* (2009) Tissue effect on genetic control of transcript isoform variation. *PLoS Genet.*, **5**, e1000608.
63. Lemos, B., Araripe, L.O., Fontanillas, P. and Hartl, D.L. (2008) Dominance and the evolutionary accumulation of cis- and trans-effects on gene expression. *Proc. Natl Acad. Sci. USA*, **105**, 14471–14476.
64. Torgerson, D.G., Boyko, A.R., Hernandez, R.D., Indap, A., Hu, X., White, T.J., Sninsky, J.J., Cargill, M., Adams, M.D., Bustamante, C.D. *et al.* (2009) Evolutionary processes acting on candidate cis-regulatory regions in humans inferred from patterns of polymorphism and divergence. *PLoS Genet.*, **5**, e1000592.
65. Emerson, J.J., Hsieh, L.C., Sung, H.M., Wang, T.Y., Huang, C.J., Lu, H.H., Lu, M.Y., Wu, S.H. and Li, W.H. (2010) Natural selection on cis and trans regulation in yeasts. *Genome Res.*, **20**, 826–836.
66. Khaitovich, P., Hellmann, I., Enard, W., Nowick, K., Leinweber, M., Franz, H., Weiss, G., Lachmann, M. and Paabo, S. (2005) Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science*, **309**, 1850–1854.
67. Blekman, R., Oshlack, A., Chabot, A.E., Smyth, G.K. and Gilad, Y. (2008) Gene regulation in primates evolves under tissue-specific selection pressures. *PLoS Genet.*, **4**, e1000271.
68. Dimas, A.S., Stranger, B.E., Beazley, C., Finn, R.D., Ingle, C.E., Forrest, M.S., Ritchie, M.E., Deloukas, P., Tavare, S. and Dermitzakis, E.T. (2008) Modifier effects between regulatory and protein-coding variation. *PLoS Genet.*, **4**, e1000244.
69. Tung, J., Fedrigo, O., Haygood, R., Mukherjee, S. and Wray, G.A. (2009) Genomic features that predict allelic imbalance in humans suggest patterns of constraint on gene expression variation. *Mol. Biol. Evol.*, **26**, 2047–2059.
70. Rifkin, S.A., Houle, D., Kim, J. and White, K.P. (2005) A mutation accumulation assay reveals a broad capacity for rapid evolution of gene expression. *Nature*, **438**, 220–223.
71. Goode, D.L., Cooper, G.M., Schmutz, J., Dickson, M., Gonzales, E., Tsai, M., Karra, K., Davydov, E., Batzoglou, S., Myers, R.M. *et al.* (2010) Evolutionary constraint facilitates interpretation of genetic variation in resequenced human genomes. *Genome Res.*, **20**, 301–310.
72. Drake, J.A., Bird, C., Nemesh, J., Thomas, D.J., Newton-Cheh, C., Raymond, A., Excoffier, L., Attar, H., Antonarakis, S.E., Dermitzakis, E.T. *et al.* (2006) Conserved noncoding sequences are selectively constrained and not mutation cold spots. *Nat. Genet.*, **38**, 223–227.
73. Athana, S., Noble, W.S., Kryukov, G., Grant, C.E., Sunyaev, S. and Stamatoyannopoulos, J.A. (2007) Widely distributed noncoding purifying selection in the human genome. *Proc. Natl Acad. Sci. USA*, **104**, 12410–12415.
74. Lomelin, D., Jorgenson, E. and Risch, N. (2010) Human genetic variation recognizes functional elements in noncoding sequence. *Genome Res.*, **20**, 311–319.
75. Barreiro, L.B., Laval, G., Quach, H., Patin, E. and Quintana-Murci, L. (2008) Natural selection has driven population differentiation in modern humans. *Nat. Genet.*, **40**, 340–345.
76. Campbell, C.D., Kirby, A., Nemesh, J., Daly, M.J. and Hirschhorn, J.N. (2008) A survey of allelic imbalance in F1 mice. *Genome Res.*, **18**, 555–563.
77. Kudaravalli, S., Veyrieras, J.B., Stranger, B.E., Dermitzakis, E.T. and Pritchard, J.K. (2009) Gene expression levels are a target of recent natural selection in the human genome. *Mol. Biol. Evol.*, **26**, 649–658.
78. Tishkoff, S.A., Reed, F.A., Ranciaro, A., Voight, B.F., Babbitt, C.C., Silverman, J.S., Powell, K., Mortensen, H.M., Hirbo, J.B., Osman, M. *et al.* (2007) Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.*, **39**, 31–40.
79. Hamblin, M.T. and Di Rienzo, A. (2000) Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. *Am. J. Hum. Genet.*, **66**, 1669–1679.
80. Tournamille, C., Colin, Y., Cartron, J.P. and Le Van Kim, C. (1995) Disruption of a GATA motif in the duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nat. Genet.*, **10**, 224–228.
81. Prabhakar, S., Visel, A., Akiyama, J.A., Shoukry, M., Lewis, K.D., Holt, A., Plajzer-Frick, I., Morrison, H., Fitzpatrick, D.R., Afzal, V. *et al.* (2008) Human-specific gain of function in a developmental enhancer. *Science*, **321**, 1346–1350.
82. Duret, L. and Galtier, N. (2009) Comment on ‘Human-specific gain of function in a developmental enhancer’. *Science*, **323**, 714.
83. Liao, B.Y. and Zhang, J. (2006) Low rates of expression profile divergence in highly expressed genes and tissue-specific genes during mammalian evolution. *Mol. Biol. Evol.*, **23**, 1119–1128.
84. Gaffney, D.J., Blekman, R. and Majewski, J. (2008) Selective constraints in experimentally defined primate regulatory regions. *PLoS Genet.*, **4**, e1000157.
85. Haygood, R., Babbitt, C.C., Fedrigo, O. and Wray, G.A. (2010) Contrasts between adaptive coding and noncoding changes during human evolution. *Proc. Natl Acad. Sci. USA*, **107**, 7853–7857.
86. Bullard, J.H., Mostovoy, Y., Dudoit, S. and Brem, R.B. (2010) Polygenic and directional regulatory evolution across pathways in saccharomyces. *Proc. Natl Acad. Sci. USA*, **107**, 5058–5063.
87. Nica, A.C., Montgomery, S.B., Dimas, A.S., Stranger, B.E., Beazley, C., Barroso, I. and Dermitzakis, E.T. (2010) Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet.*, **6**, e1000895.
88. Cookson, W., Liang, L., Abecasis, G., Moffatt, M. and Lathrop, M. (2009) Mapping complex disease traits with global gene expression. *Nat. Rev. Genet.*, **10**, 184–194.
89. Chen, Y., Zhu, J., Lum, P.Y., Yang, X., Pinto, S., MacNeil, D.J., Zhang, C., Lamb, J., Edwards, S., Sieberts, S.K. *et al.* (2008) Variations in DNA elucidate molecular networks that cause disease. *Nature*, **452**, 429–435.
90. Hsu, Y.H., Zillikens, M.C., Wilson, S.G., Farber, C.R., Demissie, S., Soranzo, N., Bianchi, E.N., Grundberg, E., Liang, L., Richards, J.B. *et al.* (2010) An integration of genome-wide association study and gene expression profiling to prioritize the discovery of novel susceptibility loci for osteoporosis-related traits. *PLoS Genet.*, **6**, e1000977.
91. Blekman, R., Man, O., Herrmann, L., Boyko, A.R., Indap, A., Kosiol, C., Bustamante, C.D., Teshima, K.M. and Przeworski, M. (2008) Natural selection on genes that underlie human disease susceptibility. *Curr. Biol.*, **18**, 883–889.
92. Bustamante, C.D., Fedel-Alon, A., Williamson, S., Nielsen, R., Hubisz, M.T., Gnanowski, S., Tanenbaum, D.M., White, T.J., Sninsky, J.J., Hernandez, R.D. *et al.* (2005) Natural selection on protein-coding genes in the human genome. *Nature*, **437**, 1153–1157.
93. Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R. *et al.* (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature*, **449**, 913–918.
94. Akey, J.M., Zhang, G., Zhang, K., Jin, L. and Shriver, M.D. (2002) Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.*, **12**, 1805–1814.